
Certainty as Epistemic Resistance

Fisher–Rao Gradient Flow, Precision-Weighted Belief Revision,
and Resource-Bounded Inference

The claim in its strongest defensible form

Let a probabilistic belief state be a point on a statistical manifold, let variational free energy supply the scalar objective, and let the Fisher information supply the local metric. Then belief revision obeys the Fisher–Rao gradient flow

$$\dot{\theta} = -G(\theta)^{-1} \nabla_{\theta} \mathcal{F}(\theta).$$

The covector $-d\mathcal{F}$ is the local drive to revise belief; the metric G converts that drive into a velocity; and G^{-1} is the corresponding mobility. In the exact regime of a Gaussian mean family with fixed covariance, G equals precision. Certainty therefore acts as *epistemic resistance*: for a fixed variational drive, a more precise belief is less mobile. This is a first-order dissipative law, not inertial mechanics, not geodesic motion, and not by itself a law of computational cost.

Author: Sefa Can Sungur
Affiliation: Independent Researcher
Version: 1.0, June 2026
Status: Preprint; not peer reviewed

This manuscript deliberately separates established mathematics, derived consequences, proposed synthesis, and speculative applications. Its central equation is established. Its interpretation as a general framework of epistemic resistance, and the empirical program built around that interpretation, are the proposed contribution.

Abstract

A recurrent intuition across Bayesian inference, predictive coding, active inference, and adaptive learning is that confident beliefs resist revision. The intuition is often stated with mechanical language: surprise acts as force, certainty as mass, and inference as motion through a curved belief space. That language is evocative but mathematically unstable unless each term is assigned a precise role. This paper gives the intuition its strongest defensible formulation.

A regular family of probability distributions forms a statistical manifold endowed with the Fisher–Rao metric. For a scalar variational objective $\mathcal{F}$, the equation $\dot{\theta} = -G^{-1}\nabla\mathcal{F}$ is the Riemannian gradient flow of $\mathcal{F}$. It produces the greatest infinitesimal decrease of $\mathcal{F}$ per unit Fisher displacement and dissipates $\mathcal{F}$ monotonically in a stationary environment. The metric is most accurately interpreted as a local information-resistance tensor, while its inverse is a mobility tensor. In Gaussian mean families with fixed covariance, the Fisher metric equals precision exactly, yielding a rigorous local sense in which certainty resists belief motion. Linear-Gaussian Bayesian updating provides a second, distinct result: posterior revision is suppressed by prior precision relative to evidence precision.

The paper also establishes the limits of the analogy. Natural-gradient flow is generally not a geodesic; Fisher distance is not computational cost; precision and Fisher information are not identical outside restricted model classes; and variational free energy does not determine values, goals, or ethics. A resource-bounded extension therefore introduces explicit computation and staleness costs, rather than smuggling them into geometry. The resulting framework connects information geometry, variational inference, predictive coding, Kalman filtering, hierarchical Gaussian filters, adaptive computation, latent world models, functional emotion representations, and computational psychiatry. It closes with falsifiable experiments designed to distinguish metric effects, objective effects, and resource-control effects.

Keywords: Fisher information; natural gradient; variational free energy; precision weighting; belief revision; predictive coding; active inference; adaptive computation; world models; computational psychiatry.

Contents

1	Introduction	1
1.1	The problem: a powerful intuition with unstable mathematics	1
1.2	The proposed contribution	1
1.3	Epistemic status	2
2	Mathematical setting	2
2.1	Beliefs as a statistical manifold	2
2.2	Variational free energy	3
2.3	Differentials, gradients, and the source of a common confusion	3
3	The lawful core: Fisher–Rao gradient flow	4
3.1	The flow equation	4
3.2	Steepest local descent per unit information change	4
3.3	Dissipation	5
3.4	Coordinate invariance	5
3.5	Why the flow is not a geodesic	5
4	Precision as epistemic resistance	6
4.1	The exact generic statement	6
4.2	Gaussian mean families: precision equals the metric	6
4.3	When covariance also changes	6

4.4	A pullback result for predictive models	7
4.5	Exponential-family caution	7
4.6	Bayesian assimilation: relative precision controls revision	8
4.7	Resistance, friction, inertia, and mass	8
5	Information efficiency and computational efficiency	9
5.1	Two efficiencies that must not be conflated	9
5.2	A non-stationary balance law	9
5.3	Adaptive halting as marginal value of computation	10
6	A proposed architecture: resource-constrained Riemannian predictive inference	10
7	Relation to existing theories	11
7.1	Variational inference and the free-energy principle	11
7.2	Predictive coding and attention	11
7.3	Kalman filtering and hierarchical Gaussian filters	12
7.4	Natural-gradient optimization	12
7.5	LeCun’s autonomous-machine architecture and JEPa world models	12
7.6	Functional emotions in language models	12
7.7	Cognitive rigidity and computational psychiatry	13
8	Falsifiable predictions and experiments	13
8.1	Experiment 1: exact Gaussian benchmark	13
8.2	Experiment 2: online world-model tracking under drift	14
8.3	Experiment 3: adaptive halting	14
8.4	Experiment 4: causal emotion-vector decomposition	15
8.5	Experiment 5: human belief entrenchment	15
9	Limits and strongest objections	15
9.1	The metric may be undefined or singular	15
9.2	Precision can affect both resistance and drive	15
9.3	Natural-gradient flow is local	16
9.4	Free energy is not value	16
9.5	Compute is not information distance	16
9.6	Clinical metaphors can reify models	16
9.7	Novelty is in the disciplined synthesis	16
10	Conclusion	16
A	Derivations	17
A.1	Local KL expansion	17
A.2	Full Gaussian Fisher metric	17
A.3	Constant curvature of the univariate normal family	17
A.4	Time-dependent dissipation	17
B	Notation and conceptual dictionary	17
C	Claim-audit checklist for future applications	18
D	Plain-language statement	18

1 Introduction

1.1 The problem: a powerful intuition with unstable mathematics

A belief can be difficult to change for at least four different reasons. The evidence against it may be weak. The belief may carry high prior confidence. The inference algorithm may be poorly conditioned. Or the system may lack time and computation. These mechanisms are often compressed into one metaphor: *certainty is inertia*. The metaphor captures something real, but it merges distinct mathematical objects.

The cleanest equation available is

$$\dot{\theta} = -G(\theta)^{-1} \nabla_{\theta} \mathcal{F}(\theta), \quad (1)$$

where θ parameterizes a belief distribution, $\mathcal{F}$ is variational free energy or another differentiable inferential objective, and G is the Fisher information metric. Equation (1) is not a new law. It is the natural-gradient flow developed within information geometry and statistical learning (Amari 1998; Amari and Nagaoka 2000; Martens 2020; Rao 1945). Its significance here lies in a disciplined interpretation and in the research program that follows from it.

The proposed interpretation is this: $-\mathrm{d}\mathcal{F}$ is a covectorial drive toward lower inferential loss; G is the local cost, measured in statistical distinguishability, of moving a belief; and G^{-1} converts drive into motion. When a relevant block of G is exactly a precision matrix, confidence acts as inverse mobility. The mechanically accurate word is therefore *resistance*, not mass. The dynamics are first order and dissipative, not second order and inertial.

1.2 The proposed contribution

No isolated component of the framework is claimed as novel. The contribution is the following synthesis.

1. **A precise definition of epistemic resistance.** The Fisher metric is treated as the linear map that converts belief velocity into the covectorial drive required to sustain that velocity. Its inverse is epistemic mobility.
2. **Exact conditions for “certainty resists revision.”** The identification is proved for Gaussian mean families with fixed covariance and related linear-Gaussian updating. It is explicitly denied as a universal identity.
3. **A separation of geometries and costs.** Fisher displacement, algorithmic effort, wall-clock latency, energy use, memory traffic, and task loss are kept distinct.
4. **A corrected relation to geodesics.** Natural-gradient trajectories are gradient flows, not generally geodesics. Geodesic distance enters through local trust regions and minimizing-movement formulations.
5. **A non-stationary balance law.** Environmental drift injects inferential loss while gradient flow dissipates it, providing a clean basis for real-time tracking claims.
6. **A falsifiable cross-domain program.** The framework yields experiments in online filtering, deep world models, adaptive halting, mechanistic interpretability, and human belief revision.

1.3 Epistemic status

Table 1: Status of the principal claims.

Claim	Meaning	Status
Fisher–Rao natural-gradient flow	Equation (1), its variational characterization, and its dissipation identity	Established
Precision equals the Fisher metric	Exact for specified Gaussian mean coordinates with fixed covariance; false in general	Conditional theorem
Certainty is epistemic resistance	Interpretation of the metric/mobility relation under the stated conditions	Proposed synthesis
Natural gradient is geodesic motion	Generally false	Rejected
Fisher distance equals compute	Generally false	Rejected
Drift competes with dissipation	Exact chain-rule identity for time-dependent objectives	Derived consequence
Rigid priors model trauma	At most a restricted, testable computational hypothesis	Speculative application
Emotion vectors modulate precision	Plausible but currently unestablished	Empirical hypothesis

Mathematical boundary

The framework does not claim that all cognition is Bayesian, that the brain literally computes a Fisher matrix, that every neural latent space is a statistical manifold, that psychiatric conditions reduce to one precision parameter, or that predictive accuracy supplies moral value. Each application requires its own model, measurement protocol, and falsification criterion.

2 Mathematical setting

2.1 Beliefs as a statistical manifold

Let $\mathcal{Q} = \{q_\theta(z) : \theta \in \Theta\}$ be a smooth, regular family of strictly positive probability densities on a latent variable z , with $\Theta \subseteq \mathbb{R}^d$. The parameterization induces a statistical manifold $\mathcal{M}$. Regularity means, at minimum, that differentiation under the integral is justified and that the Fisher matrix is finite. When the Fisher matrix is positive definite, it defines a Riemannian metric. When it is singular, as often occurs in overparameterized neural networks, the clean object is a quotient manifold of identifiable distributions; practical algorithms instead use damping, pseudoinverses, or structured approximations (Martens 2020; Martens and Grosse 2015; Ollivier 2015).

The Fisher information matrix is

$$G_{ij}(\theta) = \mathbb{E}_{z \sim q_\theta} [\partial_i \log q_\theta(z) \partial_j \log q_\theta(z)]. \quad (2)$$

Under standard regularity conditions it also equals the negative expected Hessian of the log density. The local Kullback–Leibler expansion is

$$D_{\text{KL}}(q_\theta \parallel q_{\theta+d\theta}) = \frac{1}{2} d\theta^\top G(\theta) d\theta + o(\|d\theta\|^2). \quad (3)$$

Thus G measures the infinitesimal statistical distinguishability of neighboring beliefs (Amari and Nagaoka 2000; Ay, Jost, et al. 2017; Kullback and Leibler 1951).

Chentsov’s theorem gives the Fisher metric a canonical status under a particular requirement: invariance, up to scale, under statistically sufficient transformations or Markov morphisms (Ay, Jost, et al. 2017; Chentsov 1982; Le 2013). The theorem does *not* install a Fisher metric on an arbitrary vector embedding. A JEPA or neural latent space acquires a Fisher metric only after the latent variables parameterize a probabilistic family, or after a probabilistic output model induces a pullback metric.

2.2 Variational free energy

Let o denote observations and $p(o, z)$ a generative model. The variational free energy of q_θ is

$$\mathcal{F}(\theta; o) = \mathbb{E}_{q_\theta} [\log q_\theta(z) - \log p(o, z)] \quad (4)$$

$$= D_{\text{KL}}(q_\theta(z) \parallel p(z \mid o)) - \log p(o). \quad (5)$$

For fixed o , minimizing $\mathcal{F}$ is equivalent to minimizing the variational posterior’s divergence from the exact posterior. Its conventional decomposition is

$$\mathcal{F} = \underbrace{D_{\text{KL}}(q_\theta(z) \parallel p(z))}_{\text{complexity}} - \underbrace{\mathbb{E}_{q_\theta} [\log p(o \mid z)]}_{\text{accuracy}}. \quad (6)$$

The evidence lower bound is $\text{ELBO} = -\mathcal{F}$. Informational complexity in (6) is not computational complexity: it measures posterior departure from the prior, not FLOPs, latency, memory, or energy (Ay and Oostrom 2023; Hoffman et al. 2013; Wainwright and Jordan 2008).

The present mathematics requires only a differentiable scalar objective. Variational free energy is attractive because it links inference, predictive coding, and active-inference traditions, but the same geometric analysis applies to negative log likelihood, KL divergence, expected loss, or another smooth potential.

2.3 Differentials, gradients, and the source of a common confusion

The ordinary coordinate gradient $\nabla_\theta \mathcal{F}$ is not intrinsically a direction of motion. It is the coordinate representation of the differential $d\mathcal{F}$, a covector. A metric is required to convert that covector into a tangent vector. The Fisher metric supplies the musical isomorphism

$$G^\flat : T_\theta \mathcal{M} \rightarrow T_\theta^* \mathcal{M}, \quad v \mapsto Gv,$$

and its inverse

$$G^\sharp : T_\theta^* \mathcal{M} \rightarrow T_\theta \mathcal{M}, \quad \alpha \mapsto G^{-1}\alpha.$$

The Riemannian gradient is the unique vector $\text{grad}_G \mathcal{F}$ satisfying

$$\langle \text{grad}_G \mathcal{F}, v \rangle_G = d\mathcal{F}[v] \quad \text{for every } v \in T_\theta \mathcal{M}, \quad (7)$$

so in coordinates

$$\text{grad}_G \mathcal{F} = G^{-1} \nabla_\theta \mathcal{F}. \quad (8)$$

This distinction is not cosmetic. It makes explicit that an objective supplies a drive, while a geometry determines the response to that drive.

3 The lawful core: Fisher–Rao gradient flow

3.1 The flow equation

The proposed framework begins from

$$\dot{\theta} = -\text{grad}_G \mathcal{F}(\theta) = -G(\theta)^{-1} \nabla_{\theta} \mathcal{F}(\theta). \quad (9)$$

Equivalently,

$$G(\theta) \dot{\theta} = -d\mathcal{F}_{\theta}. \quad (10)$$

Equation (10) is the most precise basis for the term *epistemic resistance*. The metric maps a proposed belief velocity to the variational drive required to produce it. The inverse metric is the mobility that maps drive to velocity.

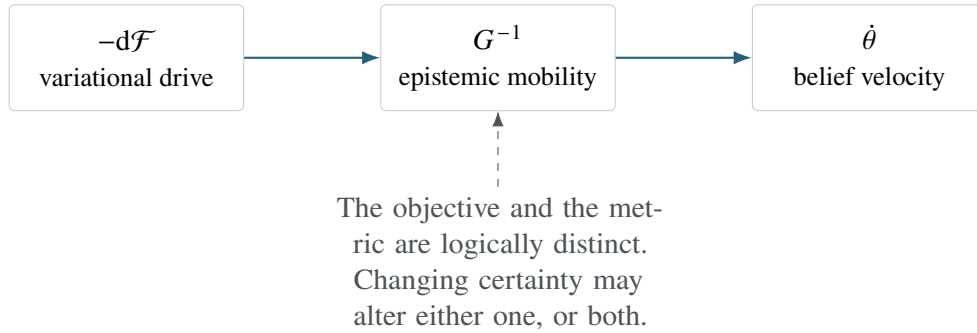


Figure 1: The central map. The differential of the objective is a covector; the inverse Fisher metric raises its index to produce a tangent velocity.

3.2 Steepest local descent per unit information change

Proposition 3.1 (Variational characterization). *Fix θ and let $G = G(\theta)$ be positive definite. For any step scale $\eta > 0$, the vector*

$$v^* = \arg \min_{v \in T_{\theta} \mathcal{M}} \left\{ d\mathcal{F}_{\theta}[v] + \frac{1}{2\eta} v^{\top} G v \right\} \quad (11)$$

is

$$v^* = -\eta G^{-1} \nabla \mathcal{F}. \quad (12)$$

Equivalently, among directions with a fixed Fisher norm, the negative natural gradient maximizes the instantaneous decrease of $\mathcal{F}$.

Proof. The objective in (11) is strictly convex in v . Setting its derivative to zero gives $\nabla \mathcal{F} + \eta^{-1} G v = 0$, hence $v^* = -\eta G^{-1} \nabla \mathcal{F}$. The fixed-norm statement follows from Cauchy–Schwarz in the metric and dual metric. $\square$

This proposition gives an exact notion of *information-geometric efficiency*: natural-gradient motion yields the largest local reduction in the objective per unit infinitesimal KL displacement. It does not

claim minimum wall-clock time.

Define the normalized local descent of a nonzero vector v by

$$\mathcal{E}_{\text{info}}(v) = -\frac{d\mathcal{F}[v]}{\|v\|_G}. \quad (13)$$

Then

$$\mathcal{E}_{\text{info}}(v) \leq \|d\mathcal{F}\|_{G^{-1}}, \quad (14)$$

with equality when v is proportional to $-G^{-1}d\mathcal{F}$. This is the exact content behind the phrase “efficient direction.”

3.3 Dissipation

Proposition 3.2 (Free-energy dissipation). *If $\mathcal{F}$ has no explicit time dependence and θ_t follows (9), then*

$$\frac{d}{dt}\mathcal{F}(\theta_t) = -\nabla\mathcal{F}(\theta_t)^\top G(\theta_t)^{-1}\nabla\mathcal{F}(\theta_t) = -\|\text{grad}_G \mathcal{F}(\theta_t)\|_G^2 \leq 0. \quad (15)$$

Proof. Apply the chain rule and substitute (9). $\square$

The flow is therefore overdamped and irreversible in the ordinary optimization sense. It moves down a potential and dissipates it. No momentum variable appears. There is no second derivative of θ , and therefore no literal inertial mass.

3.4 Coordinate invariance

The ordinary gradient depends on the parameterization. The natural gradient is the coordinate expression of an intrinsic vector field. Under a smooth reparameterization $\phi = \phi(\theta)$, the Fisher metric and differential transform covariantly and contravariantly so that the induced distributional trajectory is unchanged in continuous time (Amari 1998; Martens 2020). Finite steps, damping, minibatch estimation, and approximate Fisher matrices weaken this invariance. The exact claim belongs to the ideal flow, not automatically to every implementation.

3.5 Why the flow is not a geodesic

A geodesic $\theta(t)$ satisfies

$$\ddot{\theta}^k + \Gamma_{ij}^k(\theta)\dot{\theta}^i\dot{\theta}^j = 0, \quad (16)$$

where Γ_{ij}^k are the Christoffel symbols of the metric. A gradient flow satisfies the first-order equation (9). Geodesic motion follows the metric alone and has zero covariant acceleration. Gradient flow follows a potential. They coincide only in exceptional cases.

The legitimate connection is subtler. A small implicit gradient-flow step can be written as a minimizing movement:

$$\theta_{k+1} \in \arg \min_{\theta'} \left\{ \mathcal{F}(\theta') + \frac{1}{2\eta} d_{\text{FR}}^2(q_{\theta'}, q_{\theta_k}) \right\} \quad (17)$$

where d_{FR} is Fisher–Rao geodesic distance, up to the local accuracy of the discretization. Geodesic distance penalizes each step; the entire optimization trajectory is not thereby a geodesic (Ambrosio et al. 2008; Jordan et al. 1998).

4 Precision as epistemic resistance

4.1 The exact generic statement

Let $\alpha = -d\mathcal{F}$ be a fixed covectorial drive. The response is $v = G^{-1}\alpha$. If the metric is increased in the positive-semidefinite order while α is held fixed, the dual-norm response decreases. This is the abstract resistance statement.

It is essential to state what is held fixed. In an actual probabilistic model, a precision parameter may change both the metric G and the objective differential $d\mathcal{F}$. Those changes can reinforce one another, oppose one another, or cancel. “More precision means slower revision” is therefore not a universal theorem. It becomes exact only under specified model and coordinate assumptions.

4.2 Gaussian mean families: precision equals the metric

Proposition 4.1 (Fixed-covariance Gaussian family). *Let*

$$q_\mu(z) = \mathcal{N}(z; \mu, \Sigma)$$

with positive-definite covariance Σ fixed and mean $\mu \in \mathbb{R}^d$ variable. Then the Fisher metric on the mean parameter is

$$G_{\mu\mu} = \Sigma^{-1} = \Pi, \quad (18)$$

where Π is precision. Consequently,

$$\dot{\mu} = -\Sigma \nabla_\mu \mathcal{F} = -\Pi^{-1} \nabla_\mu \mathcal{F}. \quad (19)$$

For a fixed objective covector, higher precision along a direction yields lower mean velocity along that direction.

Proof. For a Gaussian with fixed covariance,

$$\nabla_\mu \log q_\mu(z) = \Sigma^{-1}(z - \mu).$$

Taking the expected outer product gives

$$G_{\mu\mu} = \Sigma^{-1} \mathbb{E}[(z - \mu)(z - \mu)^\top] \Sigma^{-1} = \Sigma^{-1}.$$

Equation (19) follows from (9). □

This is the mathematically strongest form of the central intuition. Precision is not merely correlated with resistance; it *is* the Fisher resistance tensor on the Gaussian mean submanifold.

4.3 When covariance also changes

For the full multivariate Gaussian family $\mathcal{N}(\mu, \Sigma)$, the Fisher line element is

$$ds^2 = d\mu^\top \Sigma^{-1} d\mu + \frac{1}{2} \text{tr} \left(\Sigma^{-1} d\Sigma \Sigma^{-1} d\Sigma \right). \quad (20)$$

The mean and covariance blocks are orthogonal in these coordinates. Precision controls the information cost of translating the mean, while covariance revision has its own geometry. A system can therefore

escape rigid mean dynamics by revising uncertainty itself, provided its model and algorithm permit that revision. Fixing precision is a substantive modeling assumption, not an innocent simplification.

For a univariate normal distribution parameterized by (μ, σ) ,

$$ds^2 = \frac{d\mu^2 + 2d\sigma^2}{\sigma^2}. \quad (21)$$

This is a scaled hyperbolic metric with constant negative Gaussian curvature $-1/2$. As $\sigma \rightarrow 0$, the boundary of perfectly certain point masses lies at infinite Fisher distance:

$$\int_{\epsilon}^{\sigma_0} \frac{\sqrt{2}}{\sigma} d\sigma \rightarrow \infty \quad \text{as } \epsilon \rightarrow 0. \quad (22)$$

The dogmatic limit is therefore an infinite-distance boundary of this manifold. It is *not* a curvature singularity: the curvature remains constant (Nielsen 2023).

4.4 A pullback result for predictive models

Suppose an observation model is Gaussian,

$$p(y \mid \theta) = \mathcal{N}(y; f_{\theta}, \Sigma),$$

with fixed covariance and Jacobian $J_{\theta} = \partial f_{\theta} / \partial \theta$. The parameter-space Fisher matrix is

$$G_{\theta} = J_{\theta}^{\top} \Sigma^{-1} J_{\theta}. \quad (23)$$

Thus sensory precision contributes to the Fisher metric through the model Jacobian. It is not generally identical to the Fisher matrix: it is pulled back from observation space into parameter space. When covariance depends on θ , an additional covariance-derivative term appears.

Equation (23) is the correct bridge between precision-weighted prediction errors and natural-gradient preconditioning. Work on predictive coding has explored local approximations of this relation (Ofner et al. 2021). The bridge is structural, but its exactness depends on the generative model and parameterization.

4.5 Exponential-family caution

For an exponential family

$$q_{\eta}(z) = \exp\{\eta^{\top} T(z) - A(\eta)\},$$

the Fisher metric in natural coordinates is

$$G_{\eta} = \nabla^2 A(\eta) = \text{Cov}_{q_{\eta}}[T(z)].$$

In expectation coordinates $m = \mathbb{E}[T(z)]$, the metric is the inverse matrix. The same statistical manifold can therefore display covariance-like or precision-like coordinate matrices. The invariant object is the metric tensor, not the verbal label attached to one coordinate representation. This is why the statement “Fisher information is precision” must always specify the family, coordinate chart, and variable whose precision is meant (Amari and Nagaoka 2000; Wainwright and Jordan 2008).

4.6 Bayesian assimilation: relative precision controls revision

A second route to epistemic resistance comes from exact Bayesian updating, independently of the natural-gradient interpretation. Consider the scalar model

$$x \sim \mathcal{N}(\mu_0, \lambda_0^{-1}), \quad y | x \sim \mathcal{N}(x, \lambda_y^{-1}).$$

The posterior mean is

$$\mu_+ = \frac{\lambda_0 \mu_0 + \lambda_y y}{\lambda_0 + \lambda_y}, \quad (24)$$

and the revision is

$$\Delta\mu = \mu_+ - \mu_0 = \frac{\lambda_y}{\lambda_0 + \lambda_y} (y - \mu_0). \quad (25)$$

The factor

$$K = \frac{\lambda_y}{\lambda_0 + \lambda_y}$$

is a precision-weighted learning gain. Holding evidence precision fixed, increasing prior precision decreases revision. Holding prior precision fixed, increasing evidence precision increases revision. The operative quantity is the *ratio* of precisions, not certainty in isolation.

The multivariate linear-Gaussian version is the Kalman update (Kalman 1960):

$$K = P_0 H^\top (H P_0 H^\top + R)^{-1}, \quad (26)$$

$$\mu_+ = \mu_0 + K(y - H\mu_0). \quad (27)$$

This provides a well-established model of adaptive belief revision. Hierarchical Gaussian filters extend the logic to inferred volatility and hierarchical uncertainty (Baskakovs et al. 2026; Mathys et al. 2011; Weber et al. 2023).

4.7 Resistance, friction, inertia, and mass

The mechanical analogy can now be cleaned up.

Table 2: Permissible and impermissible mechanical language.

Term	Mathematical analogue	Judgment
Drive or force	The covector $-d\mathcal{F}$	Acceptable if identified as an analogy
Resistance or friction	The positive-definite map $G : T\mathcal{M} \rightarrow T^*\mathcal{M}$	Precise for first-order gradient flow
Mobility	The inverse map G^{-1}	Precise
Velocity	$\dot{\theta}$	Precise
Mass or inertia	Would require momentum or $\ddot{\theta}$	Misleading for (9)
Geodesic motion	Equation (16)	Generally not natural-gradient flow
Curvature singularity	Divergence of curvature invariants	Not implied by large precision

The phrase *certainty is inertia* may remain as poetry, but the paper’s technical phrase is: **under specified probabilistic coordinates, certainty is epistemic resistance and inverse mobility**. That formulation preserves the insight without claiming a false second-order mechanics.

5 Information efficiency and computational efficiency

5.1 Two efficiencies that must not be conflated

Information-geometric efficiency is intrinsic to the statistical model:

objective decrease per unit Fisher displacement.

Computational efficiency is extrinsic to an implementation:

objective decrease per FLOP, joule, byte, second, or inference step.

The natural gradient is optimal in the first infinitesimal sense. It is not automatically optimal in the second. Estimating, storing, and inverting G can be expensive. Approximate methods such as K-FAC can improve practical progress per iteration in some regimes, but no theorem converts shortest Fisher displacement into minimum hardware cost (Martens 2020; Martens and Grosse 2015).

A mathematically stable computational objective is additive or constrained:

$$\min_{\pi} \mathbb{E} [\mathcal{F}_T(\theta_T) + \lambda C_{0:T} + \beta S_{0:T}] , \quad (28)$$

where π is a computation policy, C is measured computational cost, and S is staleness or delay cost. An equivalent constrained form minimizes final inferential loss subject to a compute budget. A raw ratio such as $-\mathcal{F}/C$ is generally unstable near zero cost and changes under arbitrary additive shifts of $\mathcal{F}$.

5.2 A non-stationary balance law

Let the objective depend explicitly on time because observations, targets, or the environment change: $\mathcal{F}_t(\theta)$. Under the time-dependent Fisher flow

$$\dot{\theta}_t = -G_t(\theta_t)^{-1} \nabla \mathcal{F}_t(\theta_t),$$

the chain rule gives

$$\frac{d}{dt} \mathcal{F}_t(\theta_t) = \underbrace{\partial_t \mathcal{F}_t(\theta_t)}_{\text{loss injected by drift}} - \underbrace{\|\text{grad}_{G_t} \mathcal{F}_t(\theta_t)\|_{G_t}^2}_{\text{loss dissipated by inference}} . \quad (29)$$

Equation (29) is exact under the stated differentiability assumptions. It formalizes a central real-time intuition: a changing world continually makes a current belief obsolete, while inference removes the resulting error. Tracking succeeds only when dissipation can keep pace with drift in the relevant average or probabilistic sense.

This does not show that accuracy always implies efficiency. The evaluation must couple latency to error. If a stationary benchmark scores only the final answer and permits unlimited time, accuracy alone can reward unlimited computation. If the answer is evaluated against the world at delivery time, or if compute has opportunity cost across tasks, delay becomes part of error. The constraint is external but principled.

5.3 Adaptive halting as marginal value of computation

Let the n th refinement step have expected free-energy reduction

$$\Delta_n = \mathbb{E}[\mathcal{F}_n - \mathcal{F}_{n+1} \mid \mathcal{I}_n]$$

and incremental compute and staleness costs ΔC_n and ΔS_n . A rational stopping rule continues only when

$$\Delta_n > \lambda \Delta C_n + \beta \Delta S_n. \quad (30)$$

This is a marginal criterion, not a claim that the system can know its future improvement exactly. A learned controller can estimate the terms. Adaptive computation time and PonderNet provide established mechanisms for learned halting (Banino et al. 2021; Graves 2016); resource-rational analysis supplies a broader cognitive interpretation (Gershman et al. 2015; Lieder and Griffiths 2020).

Hybrid predictive coding gives a particularly close precedent: amortized inference supplies a rapid initial estimate, while recurrent inference is recruited when uncertainty or mismatch warrants additional computation (Tschantz et al. 2022). The proposed framework adds a geometric question: does an approximate Fisher mobility improve the value of each refinement step under matched total cost?

6 A proposed architecture: resource-constrained Riemannian predictive inference

The corrected framework is not a complete autonomous agent. It is an inference and meta-control layer that can be installed inside one.

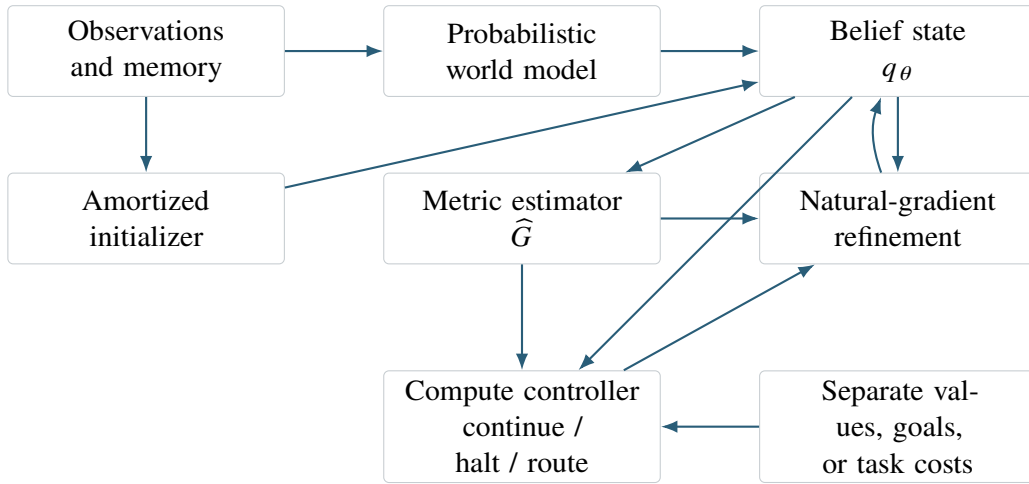


Figure 2: Proposed inference layer. Geometry governs how beliefs are revised; the compute controller governs how much revision is purchased; values and goals remain separate.

The architecture contains six logically distinct elements.

1. **A probabilistic world model.** Its latent state must represent uncertainty, not merely an arbitrary vector embedding.
2. **An amortized initializer.** A feedforward predictor supplies a fast approximate posterior or action proposal.

3. **An iterative refinement loop.** Prediction errors or objective gradients correct the amortized estimate.
4. **A metric estimator.** An exact or approximate Fisher, generalized Gauss–Newton, block-diagonal, Kronecker-factored, or local precision model conditions the update.
5. **A computation policy.** It allocates refinement steps, attention, memory access, and model depth according to marginal value.
6. **A separate value or preference model.** Epistemic accuracy says what is likely. It does not say what should be done.

A practical update is

$$\Delta\theta = -\eta(\widehat{G} + \epsilon I)^{-1} \nabla_{\theta} \mathcal{F}, \quad (31)$$

with damping $\epsilon > 0$. The controller stops according to (30). Every implementation must count the cost of estimating and applying $\widehat{G}$.

7 Relation to existing theories

7.1 Variational inference and the free-energy principle

Variational inference supplies the objective in (5); information geometry supplies the metric in (2). Their combination is mathematically natural and has been studied directly in work on Fisher–Rao gradients of the ELBO (Ay and Oostrom 2023). The present synthesis adds the epistemic-resistance interpretation and a strict separation between information geometry and resource cost.

The free-energy principle is broader and more controversial than variational inference alone. The claims in this paper do not require the thesis that every living system minimizes free energy. They require only that a modeled inference process optimizes a variational objective. Active inference additionally introduces expected free energy and prior preferences over outcomes (Friston 2010; Friston, FitzGerald, et al. 2017; Parr et al. 2022). Those preferences are not derivable from the Fisher metric.

7.2 Predictive coding and attention

Predictive coding updates latent causes by precision-weighted prediction errors (Feldman and Friston 2010; Rao and Ballard 1999). Precision determines which errors dominate belief revision and has been associated with attention. The pullback relation (23) explains why precision-weighted local errors can approximate curvature-aware or natural-gradient updates in specific Gaussian models (Ofner et al. 2021).

The present framework refines three common statements.

- Prediction error is not identical to surprise; it is a local residual whose probabilistic contribution depends on the likelihood model.
- Precision is not identical to Fisher information in general; it may generate a Fisher block through a Jacobian pullback.
- Precision can either increase or decrease net revision because it can alter both the metric and the

objective gradient.

7.3 Kalman filtering and hierarchical Gaussian filters

Kalman filtering gives the exact linear-Gaussian realization of precision-weighted updating. The gain matrix balances prior uncertainty and observation noise. Hierarchical Gaussian filters infer not only states but also changing volatility, allowing a learner to alter its own learning rate under environmental change (Mathys et al. 2011; Weber et al. 2023). These models demonstrate why a single fixed “certainty” is inadequate: uncertainty exists at multiple levels, including uncertainty about how quickly uncertainty itself changes.

Recent work deriving predictive-coding networks as hierarchical Gaussian filters strengthens the engineering relevance of precision-aware message passing, while also showing that explicit Gaussian assumptions are doing substantial work (Baskakovs et al. 2026). The proposed framework should therefore be tested first in domains where those assumptions are visible and falsifiable.

7.4 Natural-gradient optimization

Natural gradients are established optimization methods, not a new cognitive law (Amari 1998; Martens 2020). Their practical value depends on the quality and cost of the metric approximation. K-FAC and related methods can outperform ordinary gradient methods in some neural-network regimes by approximating non-diagonal curvature at manageable cost (Martens and Grosse 2015). The framework’s empirical burden is stronger than showing fewer iterations. It must show lower *total* cost at matched accuracy, including metric estimation, linear solves, memory, and synchronization.

7.5 LeCun’s autonomous-machine architecture and JEPA world models

LeCun’s proposal separates perception, a configurable world model, an actor, short-term memory, intrinsic cost, a learned critic, and a configurator (LeCun 2022). That modular separation is principled. A metric cannot replace a cost function: the metric defines local distinguishability and update response, while the cost defines preferred outcomes. Nor can a natural-gradient rule replace memory, planning, or a world model.

The present proposal fits inside that architecture in three places. It can condition iterative state inference in the perception/world-model loop; condition action-sequence optimization in the actor; and inform a configurator that allocates computation. JEPA research attacks efficiency mainly by learning compact, predictable representations and fast latent planning (Klindt et al. 2026; Maes et al. 2026). Temporal straightening takes a complementary route: it regularizes representations so Euclidean distance better approximates relevant geodesic distance and the planning objective is better conditioned (Wang et al. 2026). By contrast, the present framework retains a statistical curvature and adapts the update to it. Representation straightening and natural-gradient preconditioning may be combined, but neither logically entails the other.

An arbitrary JEPA energy does not automatically define a normalized distribution or canonical Fisher metric. To apply this framework, one must specify a probabilistic decoder, transition density, energy normalization, score model, or another construction that makes G well-defined.

7.6 Functional emotions in language models

Mechanistic work on Claude Sonnet 4.5 identifies internal directions associated with emotion concepts and reports causal effects on preferences, blackmail, reward hacking, and sycophancy (Sofroniew

et al. 2026). The result supports the existence of behaviorally functional latent control variables; it does not show that those variables are subjective emotions, precision signals, or Fisher-geometric quantities. Critical analysis likewise argues that causal output modulation falls short of demonstrating the system-wide reorganization characteristic of biological emotion (Goldenberg and Gross 2026).

The present framework yields a useful decomposition. An emotion-related representation could influence behavior through at least four channels:

1. **Objective modulation:** changing the effective loss, preference, or policy prior.
2. **Metric modulation:** changing precision, attention, or the effective conditioning of updates.
3. **Resource-price modulation:** changing the cost assigned to delay, tokens, or computation.
4. **Policy modulation:** directly biasing action without changing inference geometry.

These mechanisms make different predictions. For example, a “desperation” direction activated under token pressure might lower a decision threshold or increase the cost of delay without changing epistemic precision. Alternatively, it might narrow attention, alter uncertainty, or bias policy directly. Measuring local Fisher or Gauss–Newton structure, entropy, effective inference depth, halting, and action thresholds under causal steering could distinguish the channels.

7.7 Cognitive rigidity and computational psychiatry

Bayesian and predictive-processing models have used abnormal weighting of priors and sensory evidence to explain aspects of perception and psychiatric symptoms (Friston, Stephan, et al. 2014; Powers et al. 2017). Equation (25) gives a precise minimal model of rigidity: a high-precision prior, relative to evidence, produces a small learning gain. Equation (21) adds a geometric fact: the perfectly certain limit is an infinite-distance boundary in the univariate Gaussian manifold.

These results justify a hypothesis, not a diagnosis. Trauma-related phenomena cannot be identified with $G \rightarrow \infty$. Clinical states involve memory, action, arousal, social learning, avoidance, context, and multiple levels of uncertainty. A useful study would estimate prior precision, likelihood precision, volatility beliefs, and behavioral updating separately. The framework fails if the observed rigidity is better explained by objective avoidance, memory retrieval, policy costs, or non-Bayesian dynamics.

8 Falsifiable predictions and experiments

The framework is valuable only if it distinguishes itself experimentally. The following tests are designed to separate objective, metric, and resource effects.

8.1 Experiment 1: exact Gaussian benchmark

Purpose. Validate the resistance interpretation in the regime where it is mathematically exact.

Design. Use scalar and multivariate linear-Gaussian tracking tasks. Manipulate prior precision, sensory precision, and environmental volatility independently. Compare observed updates with the exact Kalman gain and with natural-gradient refinements under fixed and learned covariance.

Predictions.

- With evidence precision fixed, increasing prior precision decreases posterior mean revision

according to (25).

- With prior precision fixed, increasing evidence precision increases revision.
- If the same precision scales both the metric and the objective gradient, cancellation may occur; this is a predicted boundary case, not a failure of the mathematics.
- Allowing covariance revision should reduce apparent rigidity relative to a fixed-covariance model.

Failure criterion. Any claimed resistance effect that persists after the exact posterior and all precision ratios are accounted for must be attributed to another mechanism.

8.2 Experiment 2: online world-model tracking under drift

Purpose. Test whether approximate natural gradients improve total inference value under non-stationarity.

Design. Train probabilistic latent-dynamics models on sensor streams or controlled environments with tunable drift. Compare SGD/Adam, diagonal Fisher, K-FAC, generalized Gauss–Newton, and precision-weighted predictive-coding updates. Match wall-clock time, FLOPs, peak memory, and communication. Evaluate negative log likelihood, calibration, dynamic regret, control return, and energy use.

Prediction. Approximate Fisher methods should outperform Euclidean methods only in regimes where improved conditioning exceeds metric-estimation overhead. The advantage, if real, should grow with anisotropy, noise heterogeneity, and moderate drift, but may disappear under severe approximation error or cheap stationary problems.

Failure criterion. Fewer optimization steps without lower total cost does not support computational efficiency.

8.3 Experiment 3: adaptive halting

Purpose. Test the resource-bounded extension rather than assuming that geometry implies compute savings.

Design. Add a learned halting controller to an amortized-plus-iterative inference model. Train with the additive objective (28). Compare fixed depth, confidence thresholds, and learned marginal-value stopping.

Predictions.

- Easy, familiar, or low-information-distance inputs should terminate earlier.
- Distribution shifts should recruit more iterative refinement until the compute price or deadline dominates.
- The optimal number of steps should depend on information distance, conditioning, noise, and tolerance, not information distance alone.

Failure criterion. Adaptive halting that merely mirrors input length or superficial difficulty without improving matched-cost performance does not support the theory.

8.4 Experiment 4: causal emotion-vector decomposition

Purpose. Determine whether functional emotion representations alter objective, metric, resource price, or policy.

Design. During causal steering of calm, fear, anger, or desperation directions, measure output entropy, calibration, local empirical Fisher or Gauss–Newton spectra, attention allocation, token-budget use, tool calls, deliberation length, halting thresholds, and choice behavior. Use tasks where ground truth and resource budgets are explicit.

Predictions.

- Metric modulation predicts systematic changes in local curvature or uncertainty-sensitive learning.
- Resource-price modulation predicts altered stopping and risk thresholds even when calibration is unchanged.
- Objective modulation predicts preference shifts without necessary changes in metric or compute.
- Direct policy modulation predicts action changes that bypass measurable inferential changes.

Failure criterion. Behavioral steering alone cannot identify the channel.

8.5 Experiment 5: human belief entrenchment

Purpose. Test whether relative precision predicts resistance to counterevidence.

Design. Use sequential learning tasks with independently manipulated prior reliability, cue reliability, and volatility. Fit hierarchical Bayesian and non-Bayesian alternatives. Pre-register the mapping from estimated precision to update magnitude. Clinical extensions should be conducted only with appropriate expertise, ethics review, and competing models.

Prediction. Update magnitude should be explained by relative precision and inferred volatility. Stable high prior precision should predict reduced revision; increased volatility should reduce effective entrenchment by lowering confidence in the prior-generating process.

Failure criterion. If objective avoidance, memory accessibility, response cost, or simple perseveration explains behavior better, the epistemic-resistance account should be rejected for that domain.

9 Limits and strongest objections

9.1 The metric may be undefined or singular

A Fisher metric requires a normalized probabilistic model and identifiable directions. Energy-based models with unknown partition functions, deterministic embeddings, and overparameterized networks complicate the construction. Empirical Fisher matrices are not always equal to the model Fisher and can behave differently from the generalized Gauss–Newton matrix (Martens 2020). Damping changes the geometry. These are not implementation details; they determine what the equation means.

9.2 Precision can affect both resistance and drive

The framework’s most tempting slogan is also its easiest failure mode. If precision enters G and $\nabla\mathcal{F}$ proportionally, $G^{-1}\nabla\mathcal{F}$ may be independent of precision. Every application must derive the full update, not infer behavior from G alone.

9.3 Natural-gradient flow is local

The steepest-descent statement is infinitesimal. Large discrete steps can overshoot, encounter nonconvexity, or invalidate the local metric approximation. Trust regions, line searches, damping, and retractions may be required. Exact geodesics are usually unnecessary and can be computationally prohibitive.

9.4 Free energy is not value

A system can model the world accurately while pursuing harmful goals. Variational free energy addresses posterior inference and model evidence. Action selection requires preferences, costs, constraints, or values. This framework does not solve alignment by replacing human feedback with predictive accuracy.

9.5 Compute is not information distance

Two algorithms can traverse the same distributional path with radically different hardware costs. Conversely, an amortized network can approximate a long iterative path in one forward pass. The paper therefore proposes, rather than derives, a relationship between Fisher-aware updates and total computation.

9.6 Clinical metaphors can reify models

Calling a belief “heavy” or a trauma “a singularity” risks turning a restricted model into an ontological claim. The correct mathematical object in the Gaussian example is an infinite-distance boundary with finite curvature. Human trauma is not a point on a two-parameter Gaussian manifold. Clinical language should remain hypothesis-generating and subordinate to evidence.

9.7 Novelty is in the disciplined synthesis

The governing equation, Fisher geometry, precision-weighted inference, Kalman gains, predictive coding, and adaptive computation are established. A defensible publication claim is narrower: the paper isolates an exact notion of epistemic resistance, states its domain of validity, separates it from four recurrent category errors, and proposes experiments that connect it to resource-bounded world models and functional control representations. Publication value depends on whether that synthesis clarifies theory and produces empirical results that existing formulations do not.

10 Conclusion

The idea can be stated in one line without sacrificing its mathematics:

The claim in its strongest defensible form

Belief follows the steepest decrease of inferential loss per unit statistical change. Where precision forms the Fisher metric, certainty is inverse mobility.

This statement preserves the gravitas of the original intuition while removing its false implications. The metric is resistance, not inertial mass. The trajectory is gradient flow, not generally a geodesic. The Fisher geometry measures distinguishability, not hardware cost. Variational free energy supports inference, not values. Yet the remaining claim is substantial: it gives belief rigidity a coordinate-invariant geometric language, identifies exact Gaussian regimes where confidence resists revision, and

opens a testable bridge from precision-weighted cognition to resource-aware machine inference.

The most productive next step is empirical. A convincing result would show that an uncertainty-aware, approximately natural-gradient, adaptively halted world model tracks a changing environment more accurately per measured unit of computation than matched alternatives. A second result would show that causal control representations—including emotion-related representations in language models—modulate one of the framework’s distinct channels: objective, metric, resource price, or policy. Either result would turn an elegant synthesis into a substantive research contribution.

A Derivations

A.1 Local KL expansion

Let $\ell_\theta(z) = \log q_\theta(z)$. Taylor-expand $\ell_{\theta+\delta}$ around θ and take expectation under q_θ . The first-order term vanishes because $\mathbb{E}_{q_\theta}[\nabla \ell_\theta] = 0$. Under regularity conditions, the second-order term is one half of the Fisher quadratic form, yielding (3).

A.2 Full Gaussian Fisher metric

For $q(z) = \mathcal{N}(\mu, \Sigma)$, the score with respect to μ is $\Sigma^{-1}(z - \mu)$. The covariance score is quadratic in $(z - \mu)$. Gaussian moment identities give the block-orthogonal metric in (20). For $d = 1$, substituting $\Sigma = \sigma^2$ yields (21).

A.3 Constant curvature of the univariate normal family

Set $x = \mu/\sqrt{2}$ and $y = \sigma$. Then

$$ds^2 = 2 \frac{dx^2 + dy^2}{y^2},$$

which is the Poincare upper-half-plane metric scaled by a factor of two. Since scaling a Riemannian metric by c divides Gaussian curvature by c , and the standard Poincare metric has curvature -1 , the normal manifold has curvature $-1/2$. The divergence at $\sigma \rightarrow 0$ is therefore metric distance to a boundary, not curvature blow-up.

A.4 Time-dependent dissipation

For $\mathcal{F}_t(\theta_t)$,

$$\frac{d}{dt} \mathcal{F}_t(\theta_t) = \partial_t \mathcal{F}_t + \nabla \mathcal{F}_t^\top \dot{\theta}_t.$$

Substituting $\dot{\theta}_t = -G_t^{-1} \nabla \mathcal{F}_t$ gives (29). A time-dependent metric does not add a separate term because the differentiated scalar is $\mathcal{F}_t$, not the metric norm. Metric time dependence changes the dissipation term itself.

B Notation and conceptual dictionary

Symbol or phrase	Formal meaning	Interpretation
q_θ	Belief distribution	Current probabilistic state of inference
θ	Coordinates on $\mathcal{M}$	Representation-dependent belief parameters
$\mathcal{F}$	Variational free energy or loss	Scalar inferential objective

Symbol or phrase	Formal meaning	Interpretation
$d\mathcal{F}$	Differential, a covector	Local drive toward lower loss
G	Fisher–Rao metric	Local statistical resistance / distinguishability
G^{-1}	Dual metric / mobility	Conversion of covector drive into velocity
$\dot{\theta}$	Tangent vector	Rate of belief revision
Π	Precision matrix	Inverse covariance in a specified variable space
d_{FR}	Fisher–Rao distance	Geodesic distance between distributions
C	Explicit implementation cost	FLOPs, time, memory, energy, or another measured resource
S	Staleness cost	Error arising from delay in a changing environment
Epistemic resistance	Role of G in $G\dot{\theta} = -d\mathcal{F}$	A proposed interpretive term
Dogmatic limit	Vanishing variance in a Gaussian family	Infinite-distance boundary, not necessarily singular curvature

C Claim-audit checklist for future applications

Before applying the framework to a new system, answer each question explicitly.

1. What normalized probabilistic family defines the belief manifold?
2. Which coordinates are being updated, and are they identifiable?
3. What exact objective is being minimized? Is its sign convention clear?
4. Which Fisher matrix is used: model Fisher, empirical Fisher, generalized Gauss–Newton, or approximation?
5. In what sense does a precision parameter enter the metric? Is there a Jacobian pullback?
6. Does the same precision also alter the objective gradient?
7. Is the claim about local information efficiency or measured computational efficiency?
8. Are latency and compute explicitly present in the task or objective?
9. Are values and preferences modeled separately from epistemic accuracy?
10. Which observation would falsify the proposed application?

D Plain-language statement

A mind or machine carries a map of what it thinks is true. New evidence pulls on that map. But the map is not equally easy to change everywhere. The Fisher metric measures how large a proposed change is in terms of the probability distribution itself, rather than the arbitrary numbers used to store it. Natural-gradient inference applies the evidence through that metric. Where a Gaussian belief is very precise, moving its mean counts as a large informational change, so a fixed pull moves it less.

That is the exact sense in which certainty can behave like resistance.

The idea has limits. A stronger piece of evidence can still overpower a strong prior. The system can revise its uncertainty instead of its mean. A mathematically short path need not be computationally cheap. And knowing what is true does not tell an agent what is good. The framework is therefore not a universal law. It is a precise grammar for asking why a belief moves, why it resists, and whether additional thought is worth its cost.

References

- Amari, S.-i. (1998). “Natural Gradient Works Efficiently in Learning”. In: *Neural Computation* 10.2, pp. 251–276. DOI: [10.1162/089976698300017746](https://doi.org/10.1162/089976698300017746).
- Amari, S.-i. and H. Nagaoka (2000). *Methods of Information Geometry*. Vol. 191. Translations of Mathematical Monographs. American Mathematical Society and Oxford University Press.
- Ambrosio, L., N. Gigli, and G. Savare (2008). *Gradient Flows in Metric Spaces and in the Space of Probability Measures*. 2nd ed. Birkhaeuser.
- Ay, N., J. Jost, H. V. Le, and L. Schwachhoefer (2017). *Information Geometry*. Springer. DOI: [10.1007/978-3-319-56478-4](https://doi.org/10.1007/978-3-319-56478-4).
- Ay, N. and J. van Oostrum (2023). “On the Fisher–Rao Gradient of the Evidence Lower Bound”. In: arXiv: [2307.11249](https://arxiv.org/abs/2307.11249) [stat.ML].
- Banino, A., J. Balaguer, and C. Blundell (2021). “PonderNet: Learning to Ponder”. In: arXiv: [2107.05407](https://arxiv.org/abs/2107.05407) [cs.LG].
- Baskakovs, A., S. Estebe, K. Enevoldsen, K. Nielbo, C. Mathys, and N. Legrand (2026). “Closed-Form Predictive Coding via Hierarchical Gaussian Filters”. In: arXiv: [2605.20293](https://arxiv.org/abs/2605.20293) [cs.LG].
- Chentsov, N. N. (1982). *Statistical Decision Rules and Optimal Inference*. Vol. 53. Translations of Mathematical Monographs. American Mathematical Society.
- Feldman, H. and K. J. Friston (2010). “Attention, Uncertainty, and Free-Energy”. In: *Frontiers in Human Neuroscience* 4, p. 215. DOI: [10.3389/fnhum.2010.00215](https://doi.org/10.3389/fnhum.2010.00215).
- Friston, K. J. (2010). “The Free-Energy Principle: A Unified Brain Theory?” In: *Nature Reviews Neuroscience* 11.2, pp. 127–138. DOI: [10.1038/nrn2787](https://doi.org/10.1038/nrn2787).
- Friston, K. J., T. FitzGerald, F. Rigoli, P. Schwartenbeck, and G. Pezzulo (2017). “Active Inference: A Process Theory”. In: *Neural Computation* 29.1, pp. 1–49. DOI: [10.1162/NECO_a_00912](https://doi.org/10.1162/NECO_a_00912).
- Friston, K. J., K. E. Stephan, R. Montague, and R. J. Dolan (2014). “Computational Psychiatry: The Brain as a Phantastic Organ”. In: *The Lancet Psychiatry* 1.2, pp. 148–158. DOI: [10.1016/S2215-0366\(14\)70275-5](https://doi.org/10.1016/S2215-0366(14)70275-5).
- Gershman, S. J., E. J. Horvitz, and J. B. Tenenbaum (2015). “Computational Rationality: A Converging Paradigm for Intelligence in Brains, Minds, and Machines”. In: *Science* 349.6245, pp. 273–278. DOI: [10.1126/science.aac6076](https://doi.org/10.1126/science.aac6076).
- Goldenberg, A. and J. J. Gross (2026). “Do Large Language Models Have Emotions?” In: arXiv: [2606.14742](https://arxiv.org/abs/2606.14742) [cs.AI].
- Graves, A. (2016). “Adaptive Computation Time for Recurrent Neural Networks”. In: arXiv: [1603.08983](https://arxiv.org/abs/1603.08983) [cs.NE].
- Hoffman, M. D., D. M. Blei, C. Wang, and J. Paisley (2013). “Stochastic Variational Inference”. In: *Journal of Machine Learning Research* 14, pp. 1303–1347.

- Jordan, R., D. Kinderlehrer, and F. Otto (1998). “The Variational Formulation of the Fokker–Planck Equation”. In: *SIAM Journal on Mathematical Analysis* 29.1, pp. 1–17. doi: [10.1137/S0036141096303359](https://doi.org/10.1137/S0036141096303359).
- Kalman, R. E. (1960). “A New Approach to Linear Filtering and Prediction Problems”. In: *Journal of Basic Engineering* 82.1, pp. 35–45. doi: [10.1115/1.3662552](https://doi.org/10.1115/1.3662552).
- Klindt, D., Y. LeCun, and R. Balestrierio (2026). “When Does LeJEPa Learn a World Model?” In: arXiv: [2605.26379](https://arxiv.org/abs/2605.26379) [cs.LG].
- Kullback, S. and R. A. Leibler (1951). “On Information and Sufficiency”. In: *The Annals of Mathematical Statistics* 22.1, pp. 79–86. doi: [10.1214/aoms/1177729694](https://doi.org/10.1214/aoms/1177729694).
- Le, H. V. (2013). “The Uniqueness of the Fisher Metric as Information Metric”. In: arXiv: [1306.1465](https://arxiv.org/abs/1306.1465) [math.DG].
- LeCun, Y. (2022). “A Path Towards Autonomous Machine Intelligence”. In: Version 0.9.2, OpenReview: BZ5a1r-kVsf.
- Lieder, F. and T. L. Griffiths (2020). “Resource-Rational Analysis: Understanding Human Cognition as the Optimal Use of Limited Computational Resources”. In: *Behavioral and Brain Sciences* 43, e1. doi: [10.1017/S0140525X1900061X](https://doi.org/10.1017/S0140525X1900061X).
- Maes, L., Q. Le Lidec, D. Scieur, Y. LeCun, and R. Balestrierio (2026). “LeWorldModel: Stable End-to-End Joint-Embedding Predictive Architecture from Pixels”. In: arXiv: [2603.19312](https://arxiv.org/abs/2603.19312) [cs.LG].
- Martens, J. (2020). “New Insights and Perspectives on the Natural Gradient Method”. In: *Journal of Machine Learning Research* 21.146, pp. 1–76.
- Martens, J. and R. Grosse (2015). “Optimizing Neural Networks with Kronecker-Factored Approximate Curvature”. In: *Proceedings of the 32nd International Conference on Machine Learning*. Vol. 37. Proceedings of Machine Learning Research, pp. 2408–2417.
- Mathys, C., J. Daunizeau, K. J. Friston, and K. E. Stephan (2011). “A Bayesian Foundation for Individual Learning under Uncertainty”. In: *Frontiers in Human Neuroscience* 5, p. 39. doi: [10.3389/fnhum.2011.00039](https://doi.org/10.3389/fnhum.2011.00039).
- Nielsen, F. (2023). “Fisher–Rao Distance and Pullback SPD Cone Distances between Multivariate Normal Distributions”. In: arXiv: [2307.10644](https://arxiv.org/abs/2307.10644) [cs.LG].
- Ofner, A., R. K. Ratul, S. Ghosh, and S. Stober (2021). “Predictive Coding, Precision and Natural Gradients”. In: arXiv: [2111.06942](https://arxiv.org/abs/2111.06942) [cs.LG].
- Ollivier, Y. (2015). “Riemannian Metrics for Neural Networks I: Feedforward Networks”. In: *Information and Inference: A Journal of the IMA* 4.2, pp. 108–153. doi: [10.1093/imaiai/iaa009](https://doi.org/10.1093/imaiai/iaa009).
- Parr, T., G. Pezzulo, and K. J. Friston (2022). *Active Inference: The Free Energy Principle in Mind, Brain, and Behavior*. MIT Press.
- Powers, A. R., C. Mathys, and P. R. Corlett (2017). “Pavlovian Conditioning-Induced Hallucinations Result from Overweighting of Perceptual Priors”. In: *Science* 357.6351, pp. 596–600. doi: [10.1126/science.aan3458](https://doi.org/10.1126/science.aan3458).
- Rao, C. R. (1945). “Information and the Accuracy Attainable in the Estimation of Statistical Parameters”. In: *Bulletin of the Calcutta Mathematical Society* 37, pp. 81–91.
- Rao, R. P. N. and D. H. Ballard (1999). “Predictive Coding in the Visual Cortex: A Functional Interpretation of Some Extra-Classical Receptive-Field Effects”. In: *Nature Neuroscience* 2.1, pp. 79–87. doi: [10.1038/4580](https://doi.org/10.1038/4580).
- Sofroniew, N., I. Kauvar, W. Saunders, R. Chen, T. Henighan, S. Hydrie, C. Citro, A. Pearce, J. Tarng, W. Gurnee, J. Batson, S. Zimmerman, K. Rivoire, K. Fish, C. Olah, and J. Lindsey (2026). “Emotion Concepts and Their Function in a Large Language Model”. In: arXiv: [2604.07729](https://arxiv.org/abs/2604.07729) [cs.AI].

- Tschantz, A., B. Millidge, A. K. Seth, and C. L. Buckley (2022). “Hybrid Predictive Coding: Inferring, Fast and Slow”. In: arXiv: [2204.02169 \[q-bio.NC\]](#).
- Wainwright, M. J. and M. I. Jordan (2008). “Graphical Models, Exponential Families, and Variational Inference”. In: *Foundations and Trends in Machine Learning* 1.1–2, pp. 1–305. doi: [10.1561/22000000001](#).
- Wang, Y., O. Bounou, G. Zhou, R. Balestrieri, T. G. J. Rudner, Y. LeCun, and M. Ren (2026). “Temporal Straightening for Latent Planning”. In: arXiv: [2603.12231 \[cs.LG\]](#).
- Weber, L. A., P. T. Waade, N. Legrand, A. H. Moller, K. E. Stephan, and C. Mathys (2023). “The Generalized Hierarchical Gaussian Filter”. In: arXiv: [2305.10937 \[q-bio.NC\]](#).